Supplementary Materials

May 2021

Table 1: Descriptive statistics for the selected features for k-means partition

Number of PrEP interval	**N**	**%**
1	102,853	63.99%
2	38,591	24.01%
3	13,613	8.47%
4	4,088	2.54%
5	1,238	0.77%
6	268	0.17%
7	69	0.04%
8	18	0.01%
9	1	0.00%

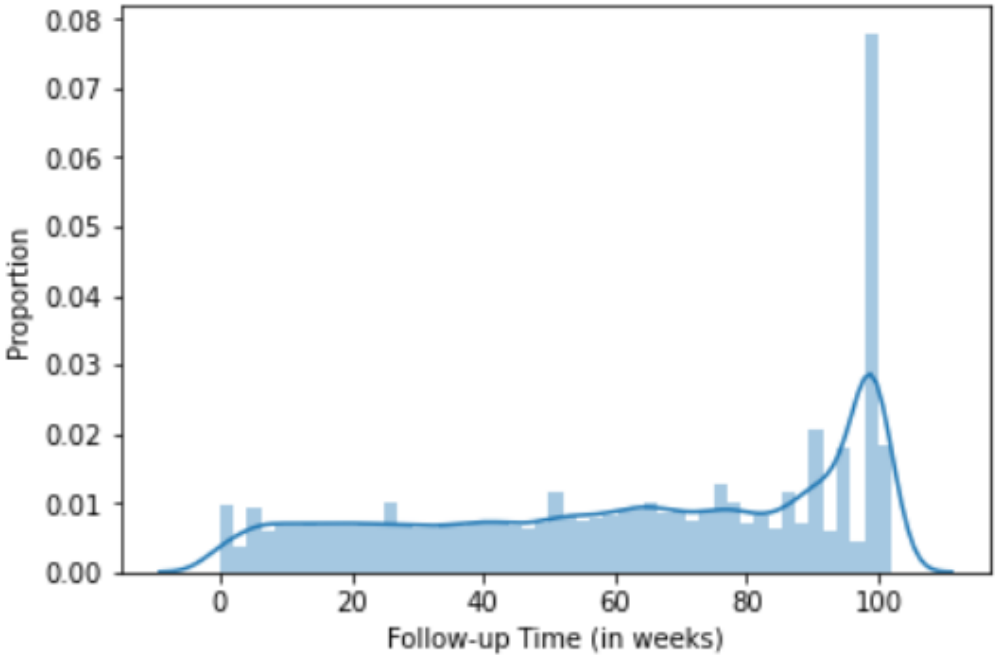

Figure 1: Distribution of the total duration in PrEP cessation

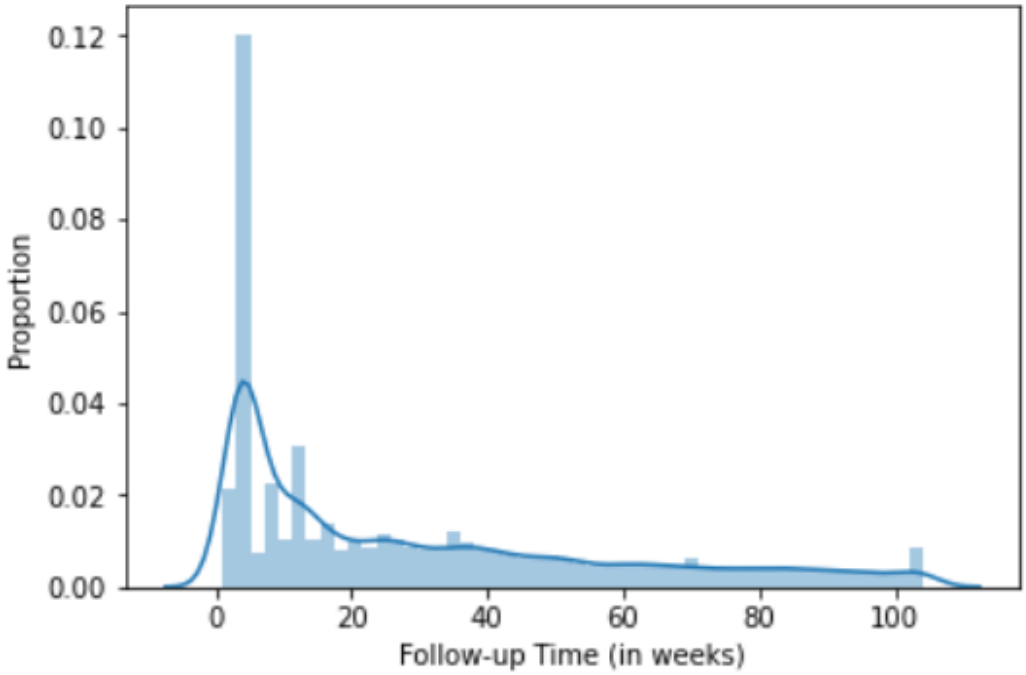

Figure 2: Distribution of the timings of the first PrEP cessation

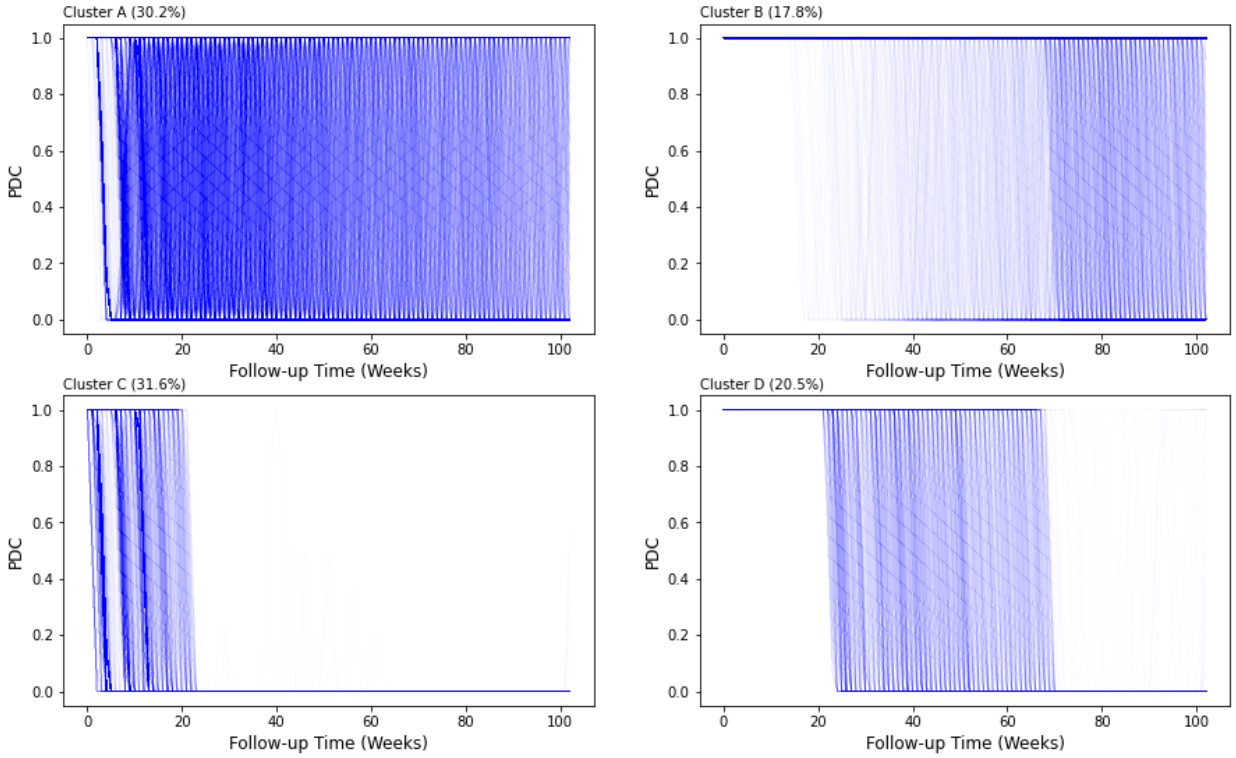

Figure 3: Distribution of the longitudinal PrEP PDC trend by stage-I cluster membership (sample size=15%)

For figure 4-7, the proportion shown on top of the median line of each stage-II cluster stands for the relative frequencies of samples assigned to the cluster.The red lines are the global average of silhouette indices under a specified k value.

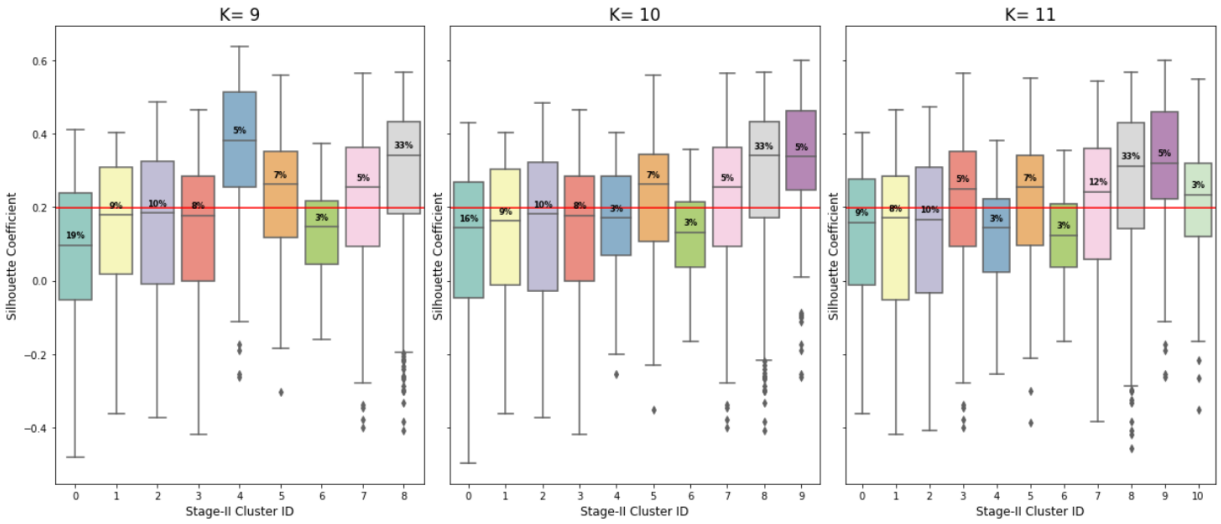

Figure 4: Distribution of silhouette indices by stage-II clusters (Stage-I Cluster A samples)

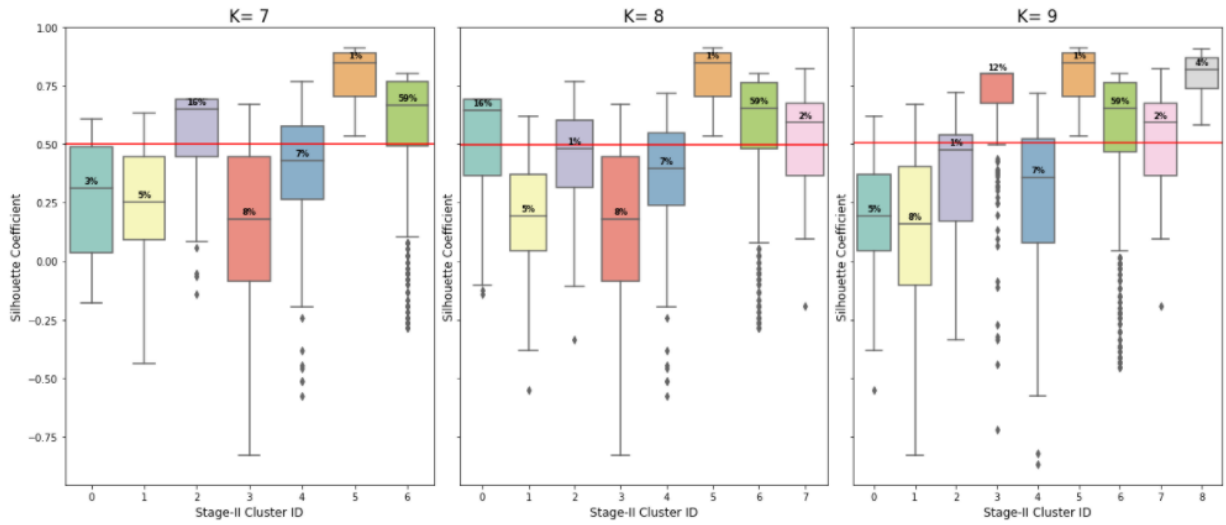

Figure 5: Distribution of silhouette indices by stage-II clusters (Stage-I Cluster B samples)

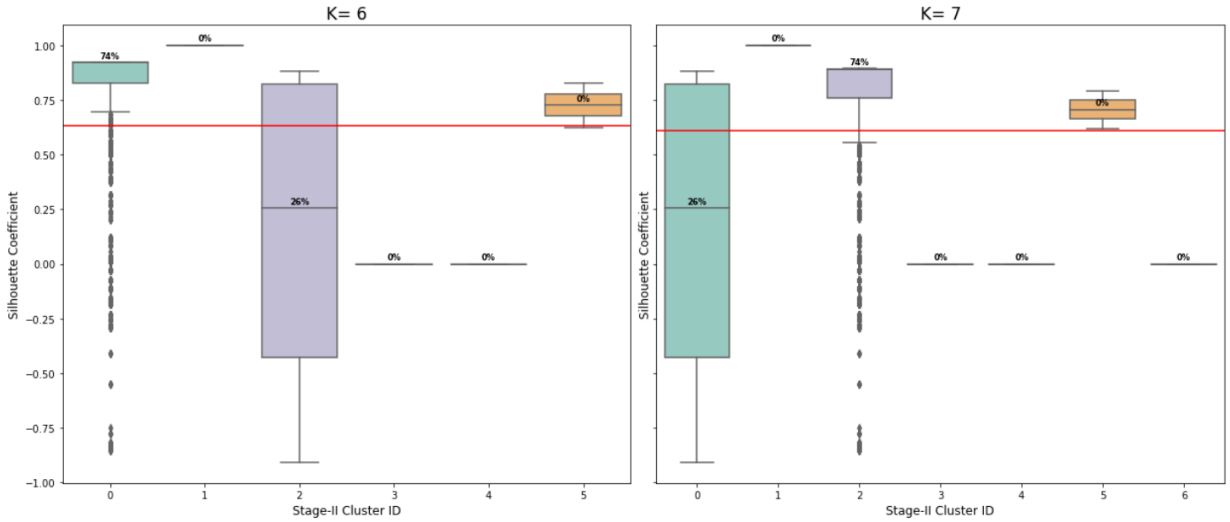

Figure 6: Distribution of silhouette indices by stage-II clusters (Stage-I Cluster C samples)

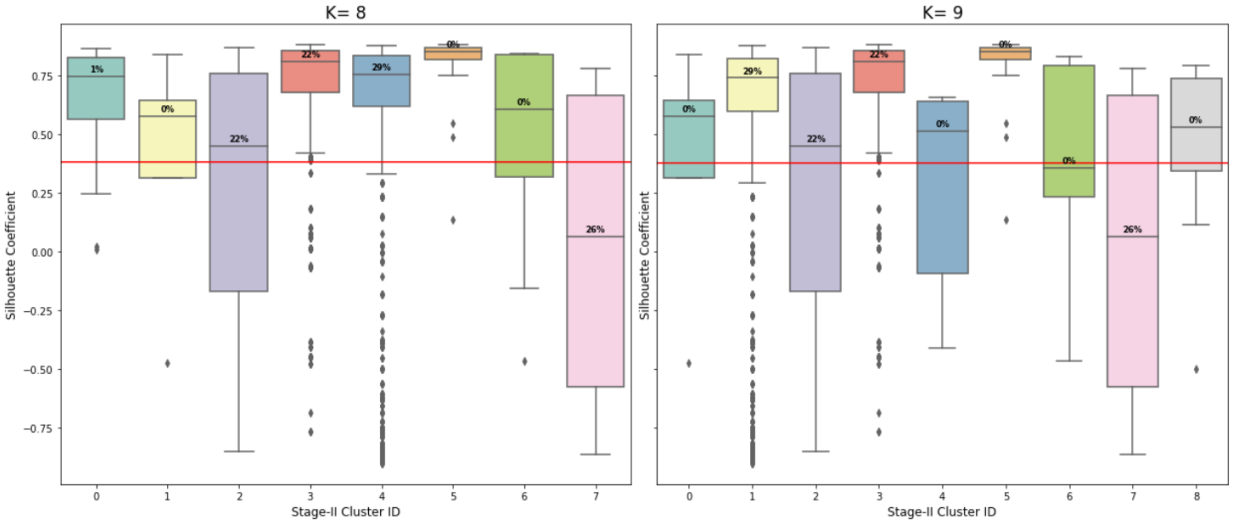

Figure 7: Distribution of silhouette indices by stage-II clusters (Stage-I Cluster D samples)

For figure 8-11, the red lines are the cluster-specific DTW DBA centroids, and the blue lines are the individual PDC time series in a cluster

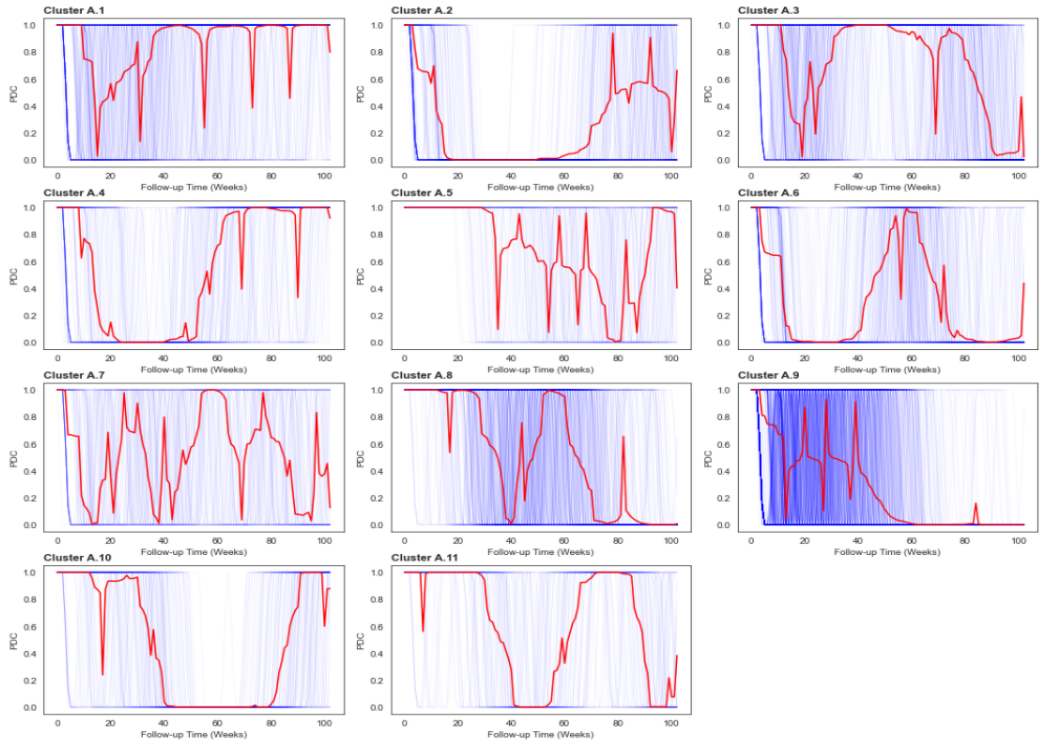

Figure 8: Distribution of longitudinal PDC trends by stage-II cluster among stage-I cluster A samples

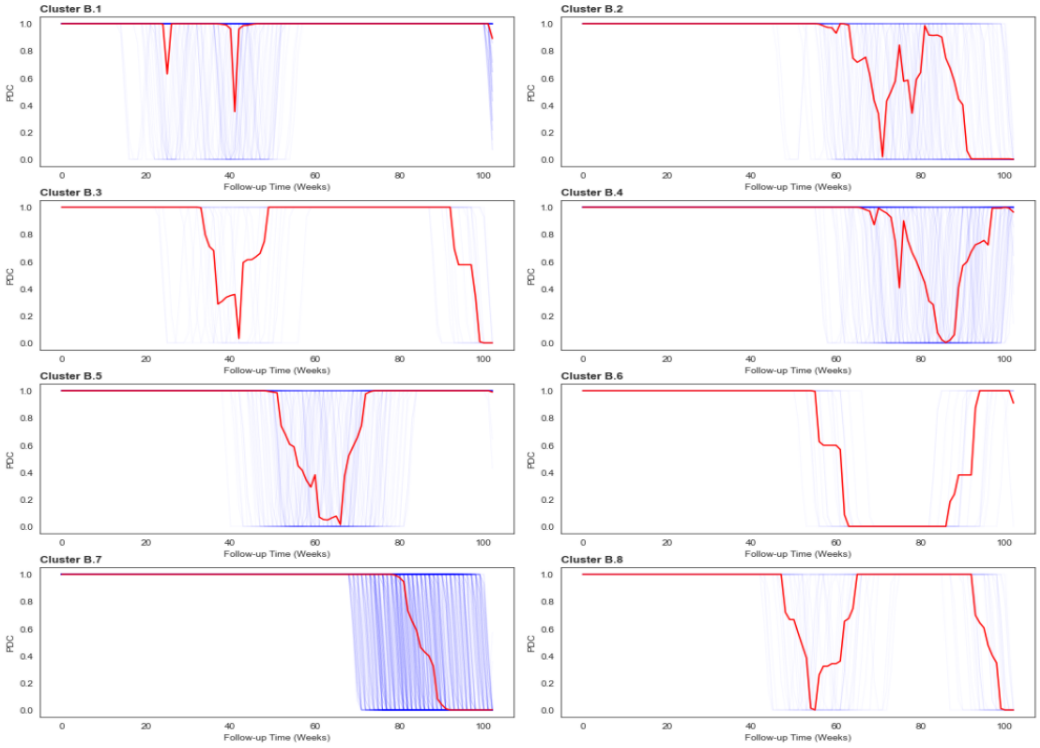

Figure 9: Distribution of longitudinal PDC trends by stage-II cluster among stage-I cluster B samples

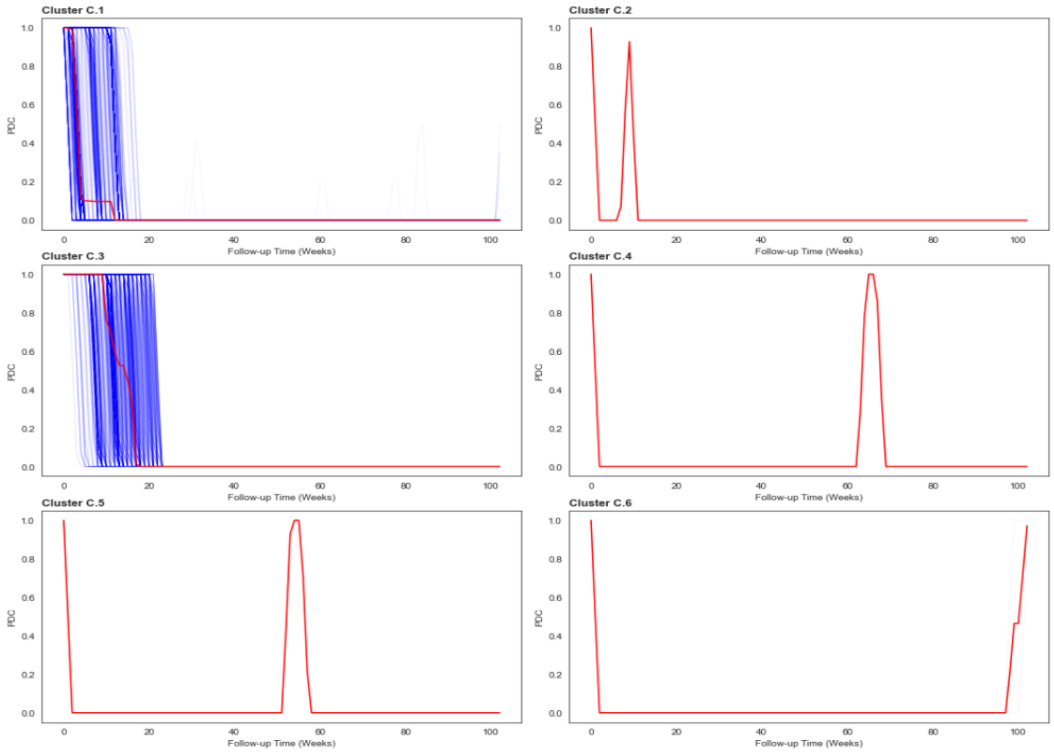

Figure 10: Distribution of longitudinal PDC trends by stage-II cluster among stage-I cluster C samples

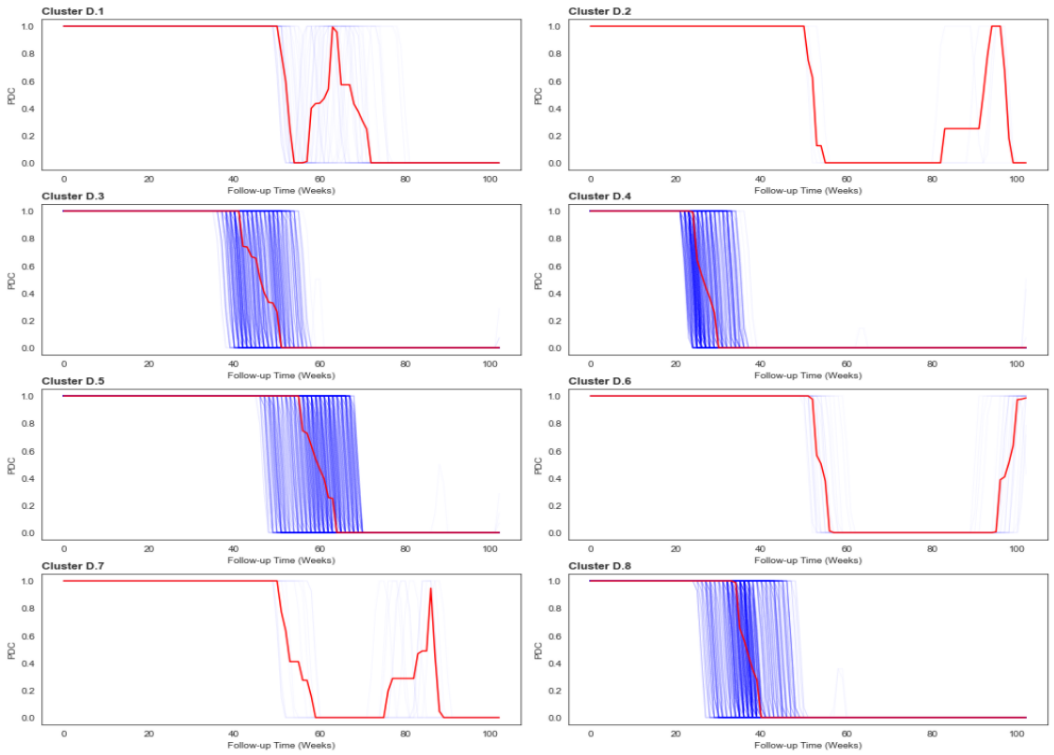

Figure 11: Distribution of longitudinal PDC trends by stage-II cluster among stage-I cluster D samples

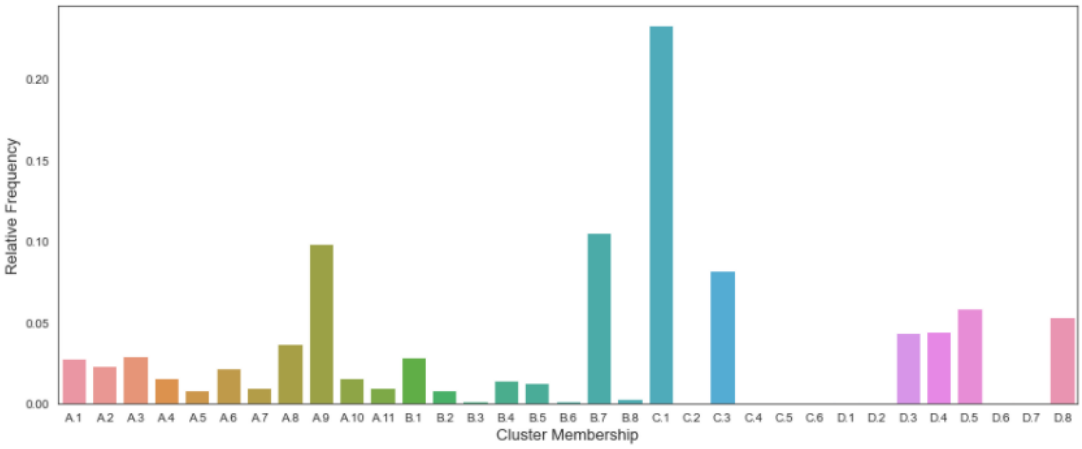

Figure 12: Marginal relative frequencies of cluster membership

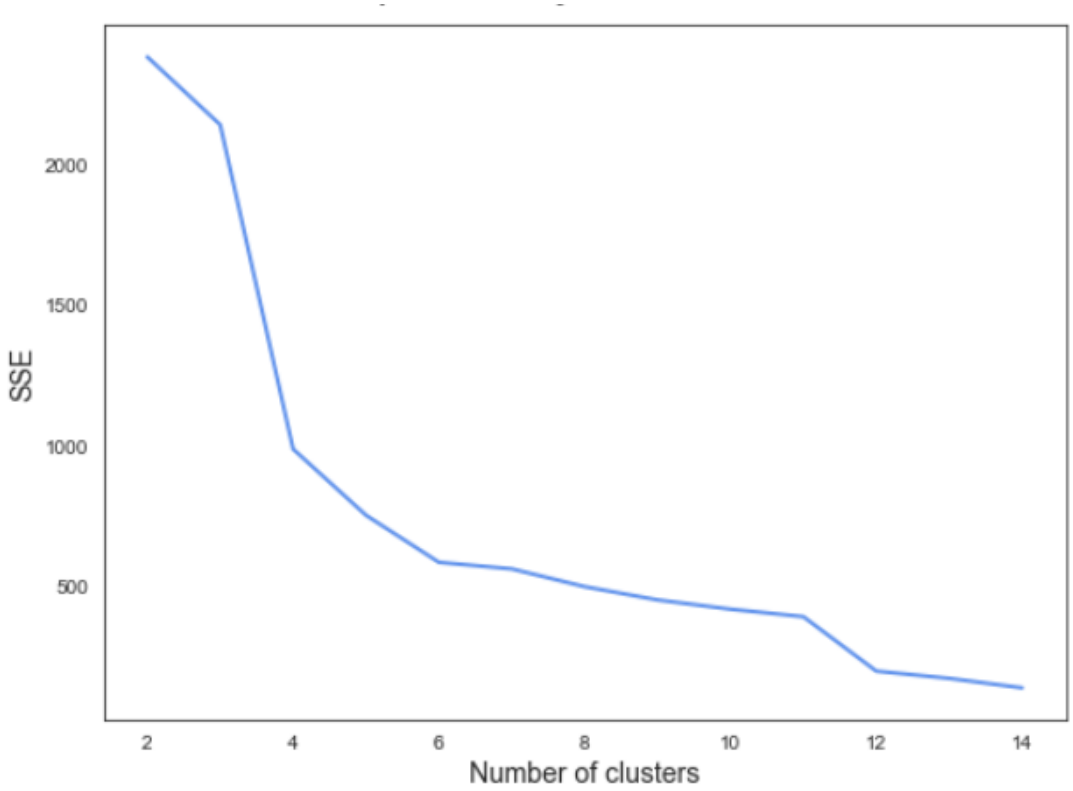

Figure 13: Sum of square error by number of DBA centroid cluster

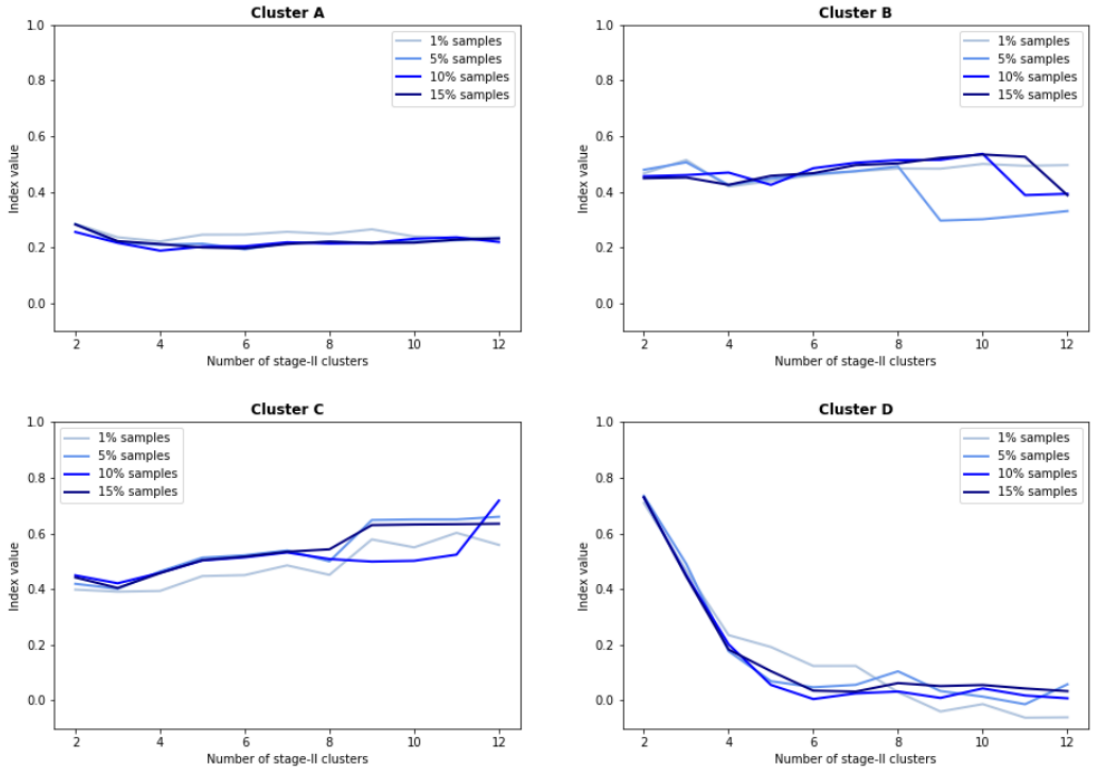

Figure 14: Average silhouette indices by number of stage-II clusters, in stage-I-clustered sample datasets (sensitivity analysis: PAM)

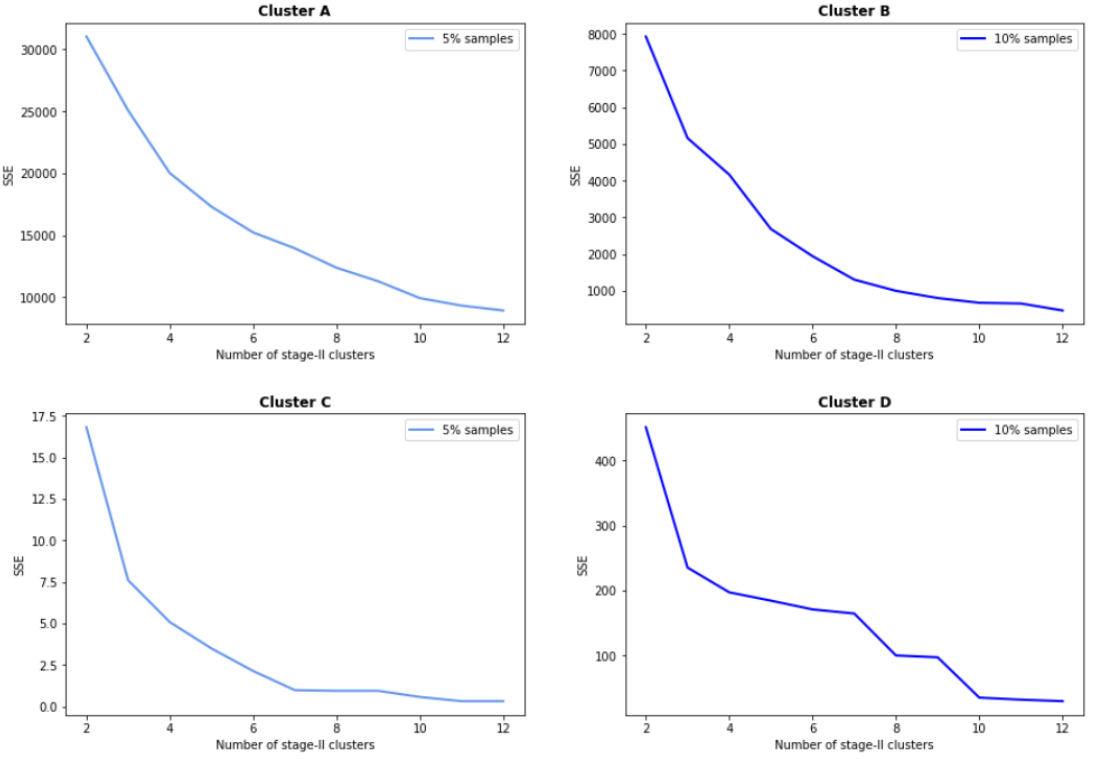

Figure 15: Sum of square error by number of stage-II clusters (sensitivity analysis: PAM)

For figure 16-19, the proportion shown on top of the median line of each stage-II cluster stands for the relative frequencies of samples assigned to the cluster. The red lines are the global average of silhouette indices under a specified k value.

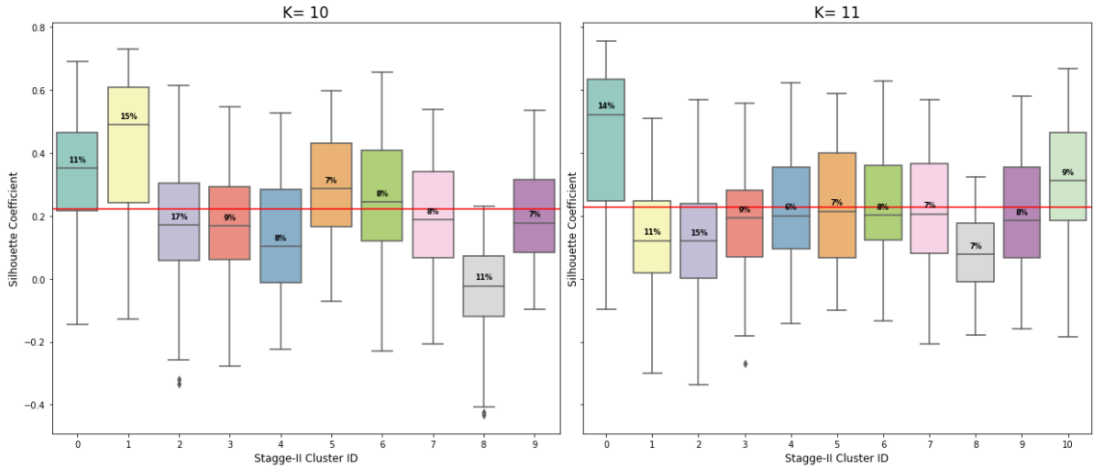

Figure 16: Distribution of silhouette indices by stage-II clusters among stage-I cluster A samples (sensitivity analysis: PAM)

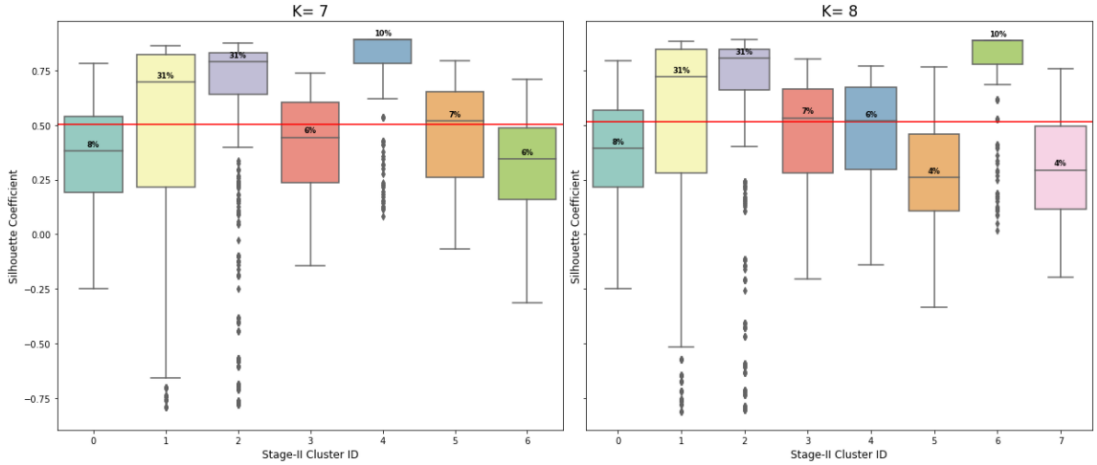

Figure 17: Distribution of silhouette indices by stage-II clusters among stage-I cluster B samples (sensitivity analysis: PAM)

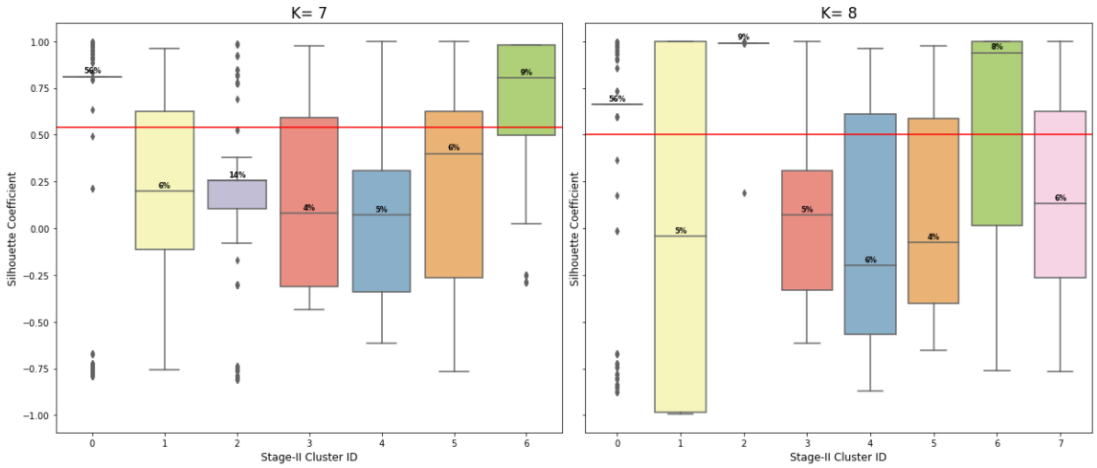

Figure 18: Distribution of silhouette indices by stage-II clusters among stage-I cluster C samples (sensitivity analysis: PAM)

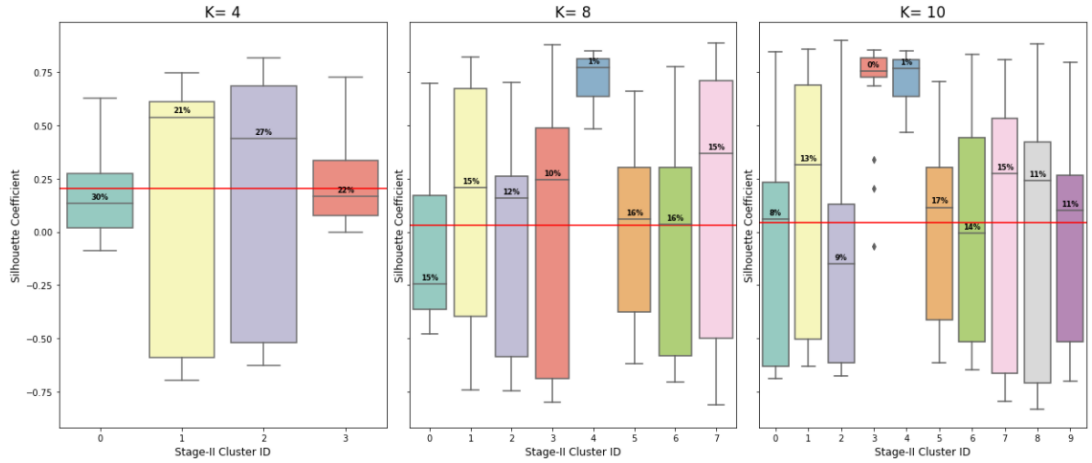

Figure 19: Distribution of silhouette indices by stage-II clusters among stage-I cluster D samples (sensitivity analysis: PAM)

For figure 20-23, the red lines are the cluster-specific medoid; the green lines are the cluster-specific DTW DBA centroids; the blue lines are the individual PDC time series in a cluster.

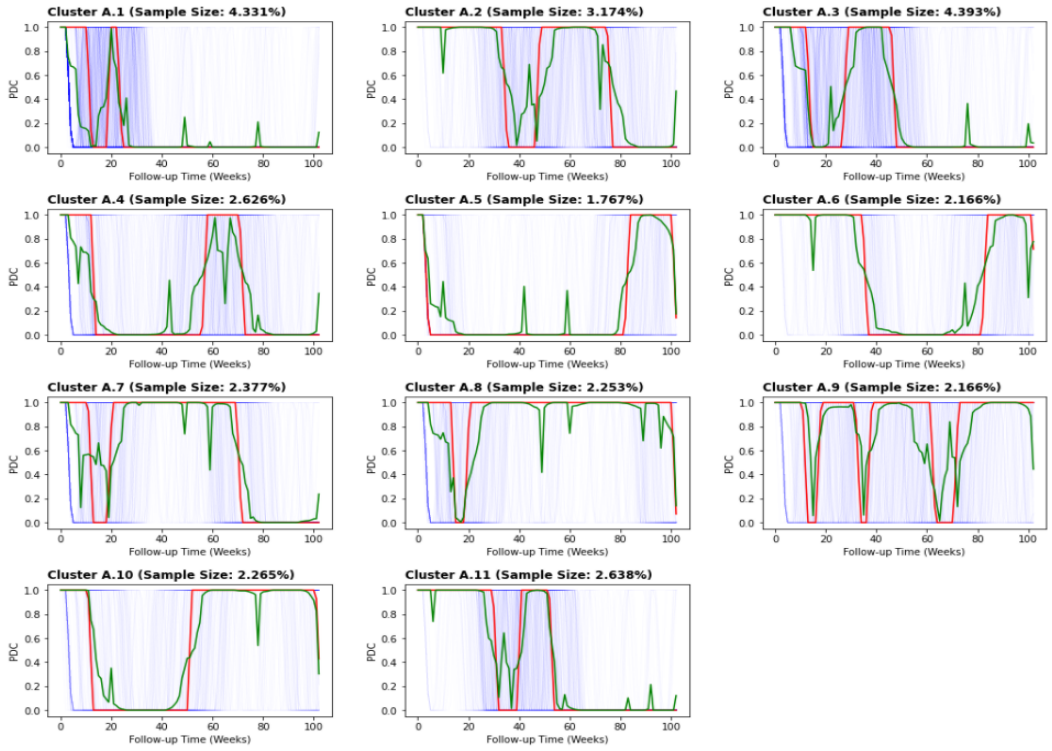

Figure 20: Distribution of longitudinal PDC trends by stage-II cluster among stage-I cluster A samples (sensitivity analysis: PAM)

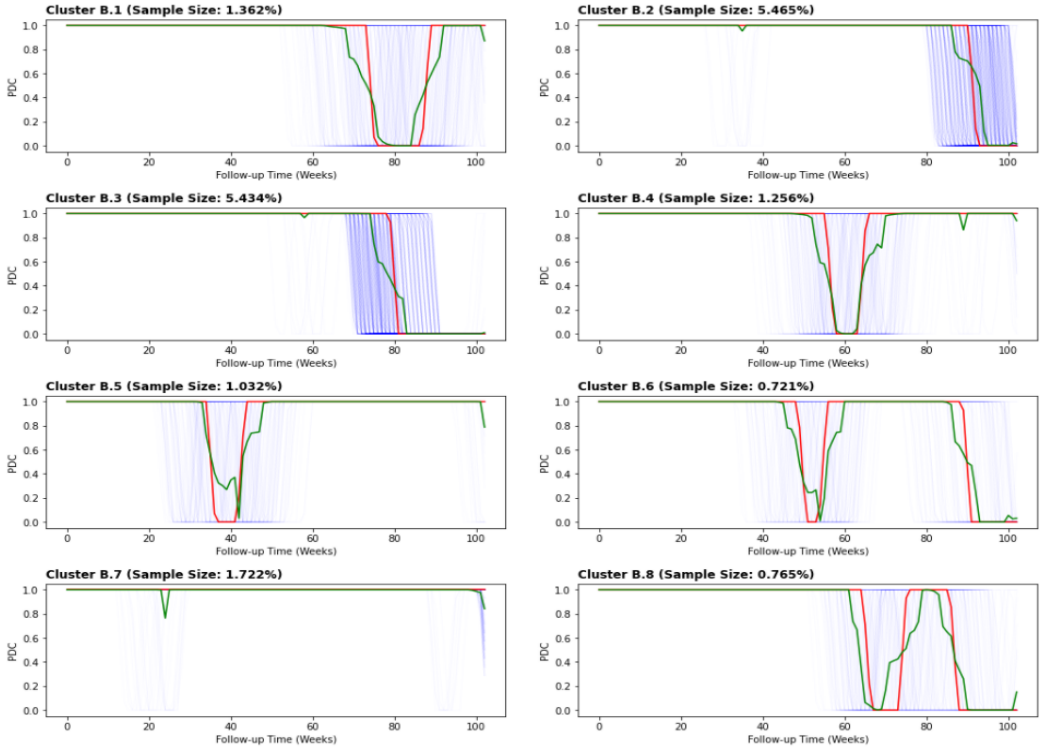

Figure 21: Distribution of longitudinal PDC trends by stage-II cluster among stage-I cluster B samples (sensitivity analysis: PAM)

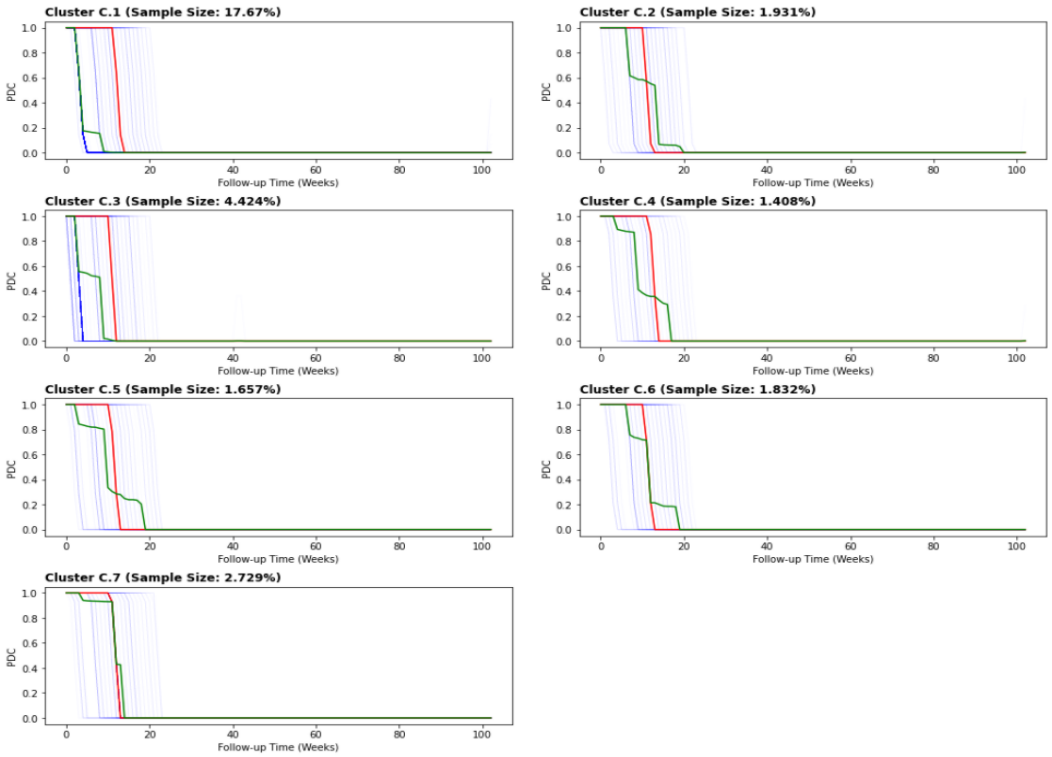

Figure 22: Distribution of longitudinal PDC trends by stage-II cluster among stage-I cluster C samples (sensitivity analysis: PAM)

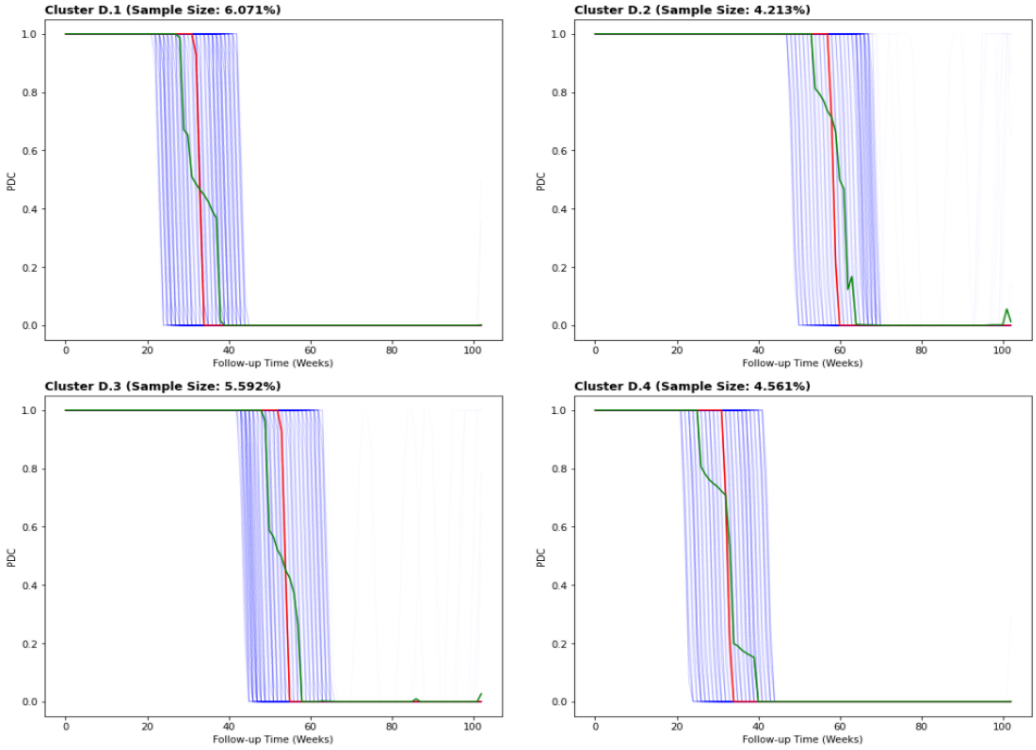

Figure 23: Distribution of longitudinal PDC trends by stage-II cluster among stage-I cluster D samples (sensitivity analysis: PAM)